Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **431,** 648–661 (2013) doi:10.1093/mnras/stt190
Advance Access publication 2013 March 07

# Connecting stellar mass and star-formation rate to dark matter halo mass out to $z \sim 2$

L. Wang,[1,2]★ D. Farrah,[1,3] S. J. Oliver,[1] A. Amblard,[4] M. Béthermin,[5,6] J. Bock,[7,8] A. Conley,[9] A. Cooray,[7,10] M. Halpern,[11] S. Heinis,[12] E. Ibar,[13] O. Ilbert,[12] R. J. Ivison,[13,14] G. Marsden,[11] I. G. Roseboom,[1,14] M. Rowan-Robinson,[15] B. Schulz,[7,16] A. J. Smith,[1] M. Viero[7] and M. Zemcov[7,8]

[1]*Astronomy Centre, Department of Physics & Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[2]*Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK*
[3]*Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA*
[4]*NASA, Ames Research Center, Moffett Field, CA 94035, USA*
[5]*Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu – CNRS – Université Paris Diderot, CE-Saclay, pt courrier 131, F-91191 Gif-sur-Yvette, France*
[6]*Institut d'Astrophysique Spatiale (IAS), bâtiment 121, Université Paris-Sud 11 and CNRS (UMR 8617), F-91405 Orsay, France*
[7]*California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA*
[8]*Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[9]*Center for Astrophysics and Space Astronomy 389-UCB, University of Colorado, Boulder, CO 80309, USA*
[10]*Department of Physics & Astronomy, University of California, Irvine, CA 92697, USA*
[11]*Department of Physics & Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada*
[12]*Laboratoire d'Astrophysique de Marseille, OAMP, Université Aix-marseille, CNRS, 38 rue Frédéric Joliot-Curie, F-13388 Marseille cedex 13, France*
[13]*UK Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[14]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[15]*Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
[16]*Infrared Processing and Analysis Center, MS 100-22, California Institute of Technology, JPL, Pasadena, CA 91125, USA*

Accepted 2013 January 31. Received 2013 January 25; in original form 2012 March 22

## ABSTRACT

We have constructed an extended halo model (EHM) which relates the total stellar mass and star-formation rate (SFR) to halo mass ($M_{\rm h}$). An empirical relation between the distribution functions of total stellar mass of galaxies and host halo mass, tuned to match the spatial density of galaxies over $0 < z < 2$ and the clustering properties at $z \sim 0$, is extended to include two different scenarios describing the variation of SFR on $M_{\rm h}$. We also present new measurements of the redshift evolution of the average SFR for star-forming galaxies of different stellar masses up to $z = 2$, using data from the *Herschel* Multi-tiered Extragalactic Survey for infrared bright galaxies.

Combining the EHM with the halo accretion histories from numerical simulations, we trace the stellar mass growth and star-formation history in haloes spanning a range of masses. We find that: (1) the intensity of the star-forming activity in haloes in the probed mass range has steadily decreased from $z \sim 2$ to 0; (2) at a given epoch, haloes in the mass range between a few times $10^{11}$ $M_\odot$ and a few times $10^{12}$ $M_\odot$ are the most efficient at hosting star formation; (3) the peak of SFR density shifts to lower mass haloes over time; and (4) galaxies that are forming stars most actively at $z \sim 2$ evolve into quiescent galaxies in today's group environments, strongly supporting previous claims that the most powerful starbursts at $z \sim 2$ are progenitors of today's elliptical galaxies.

**Key words:** methods: statistical – cosmology: observations – large-scale structure of Universe – infrared: galaxies.

## 1 INTRODUCTION

In the past 20 years or so, impressive progress has been made in characterizing the evolution of galaxy physical properties over a large fraction of cosmic time. A consistent picture, at least crudely,

★E-mail: lingyu.wang@sussex.ac.uk

has emerged in which the global stellar mass density decreases by a factor of 2 or so from $z \sim 0$ to 2 and the comoving cosmic star-formation rate (SFR) density increases by more than a factor of 10 over the past 8 Gyr, peaks around $z \sim 2$ to 3 and then declines almost linearly with time to higher redshift (e.g. Lilly et al. 1996; Madau, Pozzetti & Dickinson 1998; Dickinson et al. 2003; Rudnick et al. 2003; Schiminovich et al. 2005; Hopkins & Beacom 2006; Arnouts et al. 2007; Pascale et al. 2009; Bouwens et al. 2010). The key question that dominates both observational and theoretical efforts today is what physical processes play the dominant role in driving the evolution of the cosmic star-formation activity. Processes such as a decline in the major merger rate, reduced gas accretion in haloes, feedback from central massive black holes and supernova, and environmental effects (ram pressure stripping of gas, strangulation of the extended halo, etc.) can all impact the star-formation activity (e.g. Kereš et al. 2005, 2009; Bell et al. 2005; Bower et al. 2006; Croton et al. 2006; Lotz et al. 2008; Somerville et al. 2008).

Since galaxies form in dark matter haloes and their evolution is influenced by the accretion and successive merging of haloes (White & Rees 1978; Fall & Efstathiou 1980; Blumenthal et al. 1984), it is reasonable to assume that the physical properties of galaxies should correlate to those of the host haloes (such as the mass of the halo). Observationally, finding the mass of the host halo can be achieved in a number of ways, e.g. weak gravitational lensing (McKay et al. 2001; Hoekstra et al. 2004; Sheldon et al. 2004, 2009; Mandelbaum et al. 2006), dynamical measurement of satellite galaxies (McKay et al. 2002; Conroy et al. 2007; van den Bosch et al. 2004; More et al. 2011) and X-ray studies (Lin, Mohr & Stanford 2003; Lin & Mohr 2004; Vikhlinin et al. 2006). These techniques are at present expensive in terms of observing time and limited to low $z$ and small dynamical range in halo mass.

Alternatively, the halo model provides a simple but powerful way to statistically link galaxies with haloes. In its simplest form, the halo occupation distribution (HOD), which gives the probability of finding $N$ galaxies (with some specified properties) in a halo of mass $M_{\rm h}$, is used to interpret galaxy clustering (e.g. Peacock & Smith 2000; Seljak 2000; Scoccimarro et al. 2001; Berlind & Weinberg 2002; Zehavi et al. 2004). Modifications of the HOD include the conditional luminosity function which encodes the number of galaxies as a function of luminosity in a given halo (van den Bosch, Yang & Mo 2003; Yang et al. 2003; Vale & Ostriker 2004) and the conditional stellar mass function (CSMF) which encodes the number of galaxies as a function of stellar mass in a given halo (Yang, Mo & van den Bosch 2009; Behroozi, Conroy & Wechsler 2010; Moster et al. 2010).

In this paper, we build an extended halo model (EHM) to connect stellar mass, $m_*$, and SFR, $\psi$, with the host halo mass, $M_{\rm h}$. The EHM is a hybrid model composed of a parametrized $m_*$–$M_{\rm h}$ relation and a non-parametric $m_*$–$\psi$ relation. The first part of the EHM is to use a parametrized relation between the distribution of stellar mass and halo mass, i.e. the CSMF, to describe the statistical relation between $m_*$ and $M_{\rm h}$. The parameters in the CSMF at $z \sim 0$ are tuned by the spatial density and clustering of galaxies while their evolution in the redshift range $0 < z < 2$ is constrained by galaxy stellar mass functions (SMFs) only. The second part of the EHM is to extend the CSMF to the joint distribution in $m_*$ and $\psi$ as a function of $M_{\rm h}$, using two different scenarios for the role of $M_{\rm h}$ in determining the distribution of $\psi$ at fixed $m_*$. This second part is non-parametric as we use the observed conditional SFR distributions at fixed $m_*$ as direct inputs. The key to building the EHM is a large sample of galaxies with reliable $m_*$ and $\psi$ estimates. The *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2011) covering most of the well-studied extragalactic fields with ancillary data from the X-ray to radio is the perfect place to start such a project.

The layout of the paper is as follows. In Section 2, first we briefly describe the published measurements used to constrain the EHM. Then, we describe the data sets used to derive the stellar masses and SFRs of high-redshift galaxies in HerMES fields. In Section 3, we present the CSMF in both the local Universe and at high redshift. The evolution of the stellar content as a function of $M_{\rm h}$ is derived using the CSMF as a function of redshift and the halo accretion history from $N$-body simulations. In Section 4, we extend the CSMF to the 2D distribution of galaxies in the ($\psi$, $m_*$) plane as a function of $M_{\rm h}$. The evolution of the star-formation activity as a function of $M_{\rm h}$ is derived using the EHM as a function of redshift and the halo accretion history. Finally, conclusions and discussions are presented in Section 5. Unless otherwise stated, we assume $\Omega_{\rm M} = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.8$ and $h = 0.7$. All magnitudes are in the AB system.

## 2 DATA SETS

To constrain the $m_*$–$M_{\rm h}$ relation at $z \sim 0$, we use the published SMF (Guo et al. 2010) and correlation functions of the Sloan Digital Sky Survey (SDSS) galaxies (Li et al. 2006). To constrain the redshift evolution of the $m_*$–$M_{\rm h}$ relation, we use the published SMFs in Pérez-González et al. (2008) based on a combined sample of 3.6 and 4.5 μm selected sources in the *Hubble Deep Field*-North (HDF-N), the *Chandra Deep Field*-South (CDF-S) and the Lockman Hole. The clustering properties of high-redshift galaxies are not used to constrain the evolution parameters in the $m_*$–$M_{\rm h}$ relation due to issues explained in Section 3.2.

To extend the CSMF to the joint distribution in $m_*$ and $\psi$ as a function of $M_{\rm h}$, we use the conditional probability distribution function (PDF) of SFR of the entire population as a function of $m_*$. The conditional PDF of SFR of galaxies in the local Universe is taken from Salim et al. (2007). To derive the conditional SFR distribution as a function of $m_*$ in the distant Universe, we use galaxies observed in three well-studied extragalactic fields, the Extended *Chandra Deep Field*-South (ECDFS) field, the Cosmic Evolution Survey (COSMOS) field and the Extended Groth Strip (EGS). We will describe in detail the data sets used in each field below.

### 2.1 COSMOS

The COSMOS photometric redshift catalogue derived from broad and medium bands [*GALEX* far-ultraviolet (FUV) and near-ultraviolet (NUV), optical to infrared data $u^*B_JV_Jg^+r^+i^+i^*z^+JK_sK$, 14 medium and narrow bands from Subaru and 4 Infrared Array Camera (IRAC) channels] is described in Ilbert et al. (2009). We use an updated version (v1.8 dated from 2010 July 13) of Ilbert et al. (2009). The quality of the photometric redshift is very high with $1\sigma$ in $(1 + z) \sim 0.007$ at $i^+_{\rm AB} < 22.5$. At $i^+_{\rm AB} < 24$ and $z < 1.25$, $1\sigma$ in $(1 + z) \sim 0.012$. The deep NIR and IRAC coverage enables the photo-$z$ to be extended to $z \sim 2$, with $1\sigma$ in $(1 + z) \sim 0.06$ at $i^+_{\rm AB} \sim 24$. Following Ilbert et al. (2010), we construct a mass-selected sample as generated from the 3.6 μm catalogue of the S-COSMOS survey (Sanders et al. 2007). We cross-match the 3.6 μm and the latest photo-$z$ catalogue by taking the nearest match within 1 arcsec. The probability of incorrect identification is <1 per cent (Ilbert et al. 2010). We then select sources with $f_{3.6} \geq 5$ μJy (the 90 per cent complete limit), around 2.8 per cent of which are

not matched to an optical counterpart. Using the public photo-$z$ catalogue from the NEWFIRM Medium-Band Survey (Whitaker et al. 2011) covering a small area of the COSMOS field but deeper than Ilbert et al. (2009), we estimate that only 1 per cent (2.5 per cent) of the sources with $f3.6 \geq 5$ μJy lie at $z < 1.6$ ($z < 2.0$) or do not have a photo-$z$ estimate. As we are only concerned with the relation between $m_*$, $\psi$ and $M_{\rm h}$ at $z < 2$, we will ignore this 1 per cent of 3.6 μm sources in our analysis.

### 2.2 ECDFS

We use the Multiwavelength Survey by Yale–Chile (MUSYC) Subaru v1.0 Catalog (Cardamone et al. 2010) containing over 84 400 sources. The catalogue includes photometry in 32 MUSYC images of the ECDF-S region, including optical to infrared data (*UU38BVRIzJHK*), 18 medium bands from Subaru and 4 IRAC channels as part of the SIMPLE survey (Damen et al. 2011), for all sources detected in the combined *BVR* image. Photometric redshifts are determined using the EASY code (Brammer, van Dokkum & Coppi 2008). The quality of the photometric redshifts is very high, with $1\sigma = 0.007$ in $(1 + z)$ in the $z = [0.1, 1.2]$, similar to that of the COSMOS field. At $z = [1.2, 3.7]$, the photometric redshift accuracy gets worse with $1\sigma = 0.02$ in $(1 + z)$. We select 3.6 μm sources above the completeness limit which is 1 μJy.

### 2.3 EGS

We use a 3.6 + 4.5 μm selected catalogue in the Extended Groth Strip (EGS) containing 28-band photometry from the ultraviolet to the far-infrared (*GALEX* FUV and NUV, CFHTLS $u^*g'r'i'z'$, MMT $u'giz$, CFHT12k BRI, ACS $V_{606}i_{814}$, Subaru R, NICMOS $J_{110}H_{160}$, MOIRCS $K_s$, CAHA $JK_s$, WIRC *JK* and 4 IRAC channels) (Barro et al. 2011a,b). The typical photometric redshift accuracy is $1\sigma = 0.034$ in $(1 + z)$, with a catastrophic outlier fraction of 2 per cent. We apply the 90 per cent completeness limit at 3.6 μm by selecting sources with $f3.6 \leq 2.3$ μJy over areas with homogeneous depth $52.^\circ025 \leq \delta \leq 53.^\circ525$. We also mask out regions in the wings of bright stars.

### 2.4 Deriving stellar mass and SFR from HerMES and ancillary data

We use the LE PHARE code (Arnouts et al. 2002; Ilbert et al. 2006) and the Bruzual & Charlot (2003) stellar population synthesis models to derive stellar properties such as stellar mass and SFR. We use the same parameters as in Ilbert et al. (2010) to generate the spectral energy distribution (SED) templates, e.g. a Chabrier initial mass function (Chabrier 2003), two different metallicities (solar and subsolar) and an exponentially declining star-formation history. Dust extinction is applied to the templates using the Calzetti et al. (2000) law.

We cross-match the 3.6 μm catalogue in each field with the 24 μm catalogue by taking the nearest match within 2 arcsec. The SPIRE[1] fluxes of the 24 μm sources are obtained using a combination of linear inversion and model selection technique (Roseboom et al. 2010; Rosebomm et al. 2012). With SPIRE, we are able to probe the rest-frame far-IR region to constrain the infrared luminosity $L_{\rm IR}$ (integrated from 8 to 1000 μm). Previous studies extrapolate $L_{\rm IR}$ from the 24 μm data and the resulting $L_{\rm IR}$ can be overestimated by a factor of 5 at $z > 1.5$ (Daddi et al. 2007; Papovich et al. 2007; Murphy et al. 2009; Elbaz et al. 2010; Nordon et al. 2012). We use the Chary & Elbaz (2001) templates to fit the infrared SEDs of galaxies observed at 24 μm and at least one SPIRE band to calculate $\psi_{\rm IR} = 1.09 \times 10^{-10} \times L_{\rm IR}$ (Kennicutt 1998). For galaxies not observed in any SPIRE band (around 70 per cent of the 3.6 μm selected sample), we use $\psi_{\rm SED}$ derived from SED fitting to the UV to MIR photometric data.

In each field, we generate 10 Monte Carlo realizations of the original photo-$z$ catalogue using the redshift PDF of each galaxy and repeat the stellar mass and SFR calculation. In Fig. 1, we plot the conditional SFR distributions as a function of $m_*$ in six redshift bins, $z = [0.2, 0.5]$, $[0.5, 0.8]$, $[0.8, 1.0]$, $[1.0, 1.3]$, $[1.3, 1.6]$ and $[1.6, 2.0]$, averaged over all Monte Carlo realizations in COSMOS, ECDFS and EGS. The star-forming sequence[2] can be clearly seen and it evolves upwards roughly independently of $m_*$. The number of quiescent massive galaxies gradually builds up as redshift decreases. In each redshift bin, the conditional SFR distribution in a given stellar mass bin can be modelled as the sum of two Gaussian distributions which represent the star-forming and passive populations

$$\Phi(\psi|m_*) = \Phi_{\rm star-forming}(\psi|m_*) + \Phi_{\rm passive}(\psi|m_*). \quad (1)$$

In this paper, we define star-forming galaxies as those with SFR $\geq \langle\psi\rangle_{\rm star-forming} - 2\sigma_{\rm star-forming}$, where $\langle\psi\rangle_{\rm star-forming}$ and $\sigma_{\rm star-forming}$ are the mean SFR and standard deviation of the star-forming population, respectively. The advantage of our definition of star-forming galaxies is that it naturally takes into account the fact that the SFR of a star-forming galaxy increases with increasing stellar mass and increasing redshift (as shown in Fig. 1). In Fig. 2, we plot the redshift evolution of the average SFR as a function of stellar mass for star-forming galaxies and the best-fitting power law to points above the stellar mass completeness limit in each redshift slice (see Table 1). The best-fitting parameters in the power-law fitting of the $m_*-\psi$ relation are listed in Table 2.

## 3 EHM: 1. CONNECTING STELLAR MASS WITH HALO MASS

### 3.1 The stellar-to-halo mass relation at $z \sim 0$

We choose the CSMF, $\Phi(m_*|M_{\rm h})$, which specifies the number of galaxies of stellar mass $m_*$ that reside in a halo of mass $M_{\rm h}$, to describe the stellar-to-halo mass relation. Details of the parametrization of the CSMF and the fitting process to the observed spatial density and clustering of galaxies can be found in Appendix A. The left-hand panel in Fig. 3 compares the measured SMF of the local Universe (Guo et al. 2010) with the best-fitting SMF from our CSMF. Note that in comparing to the observed SMF, the predicted SMF from the CSMF (the black dashed line in the left-hand panel of Fig. 3) has been convolved with a log-normal distribution with its width set to 0.1 dex (Li & White 2009) to account for statistical errors in the observational estimate of stellar mass. It is clear that central galaxies dominate the SMF over the entire mass range

[1] The Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) is one of three scientific instruments onboard *Herschel* (Pilbratt et al. 2010). It operates in three wavelength bands centred at 250, 350 and 500 μm.

[2] For star-forming galaxies, there exists a strong correlation between stellar mass $m_*$ and SFR $\psi$ (with an estimated intrinsic scatter of ∼0.3 dex) from $z \sim 0$ to 3 (e.g. Daddi et al. 2007; Elbaz et al. 2007; Noeske et al. 2007; Rodighiero et al. 2010; Karim et al. 2011).

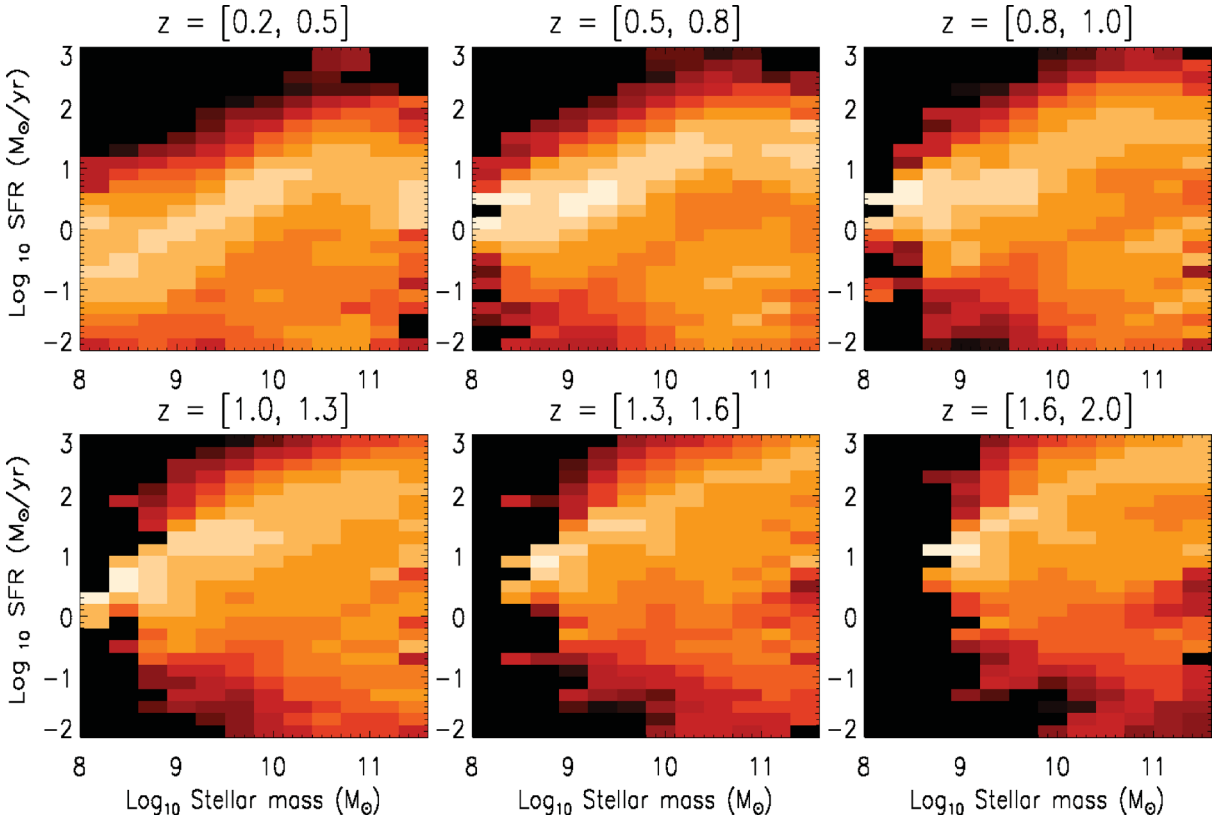


**Figure 1.** The conditional PDF of SFR as a function of stellar mass, in six redshift bins (from left to right and top to bottom, $z$ = [0.2, 0.5], [0.5, 0.8], [0.8, 1.0], [1.0, 1.3], [1.3, 1.6] and [1.6, 2.0]), averaged over COSMOS, ECDFS and EGS. The star-forming sequence can be clearly seen in each panel and it evolves upwards roughly independently of stellar mass.

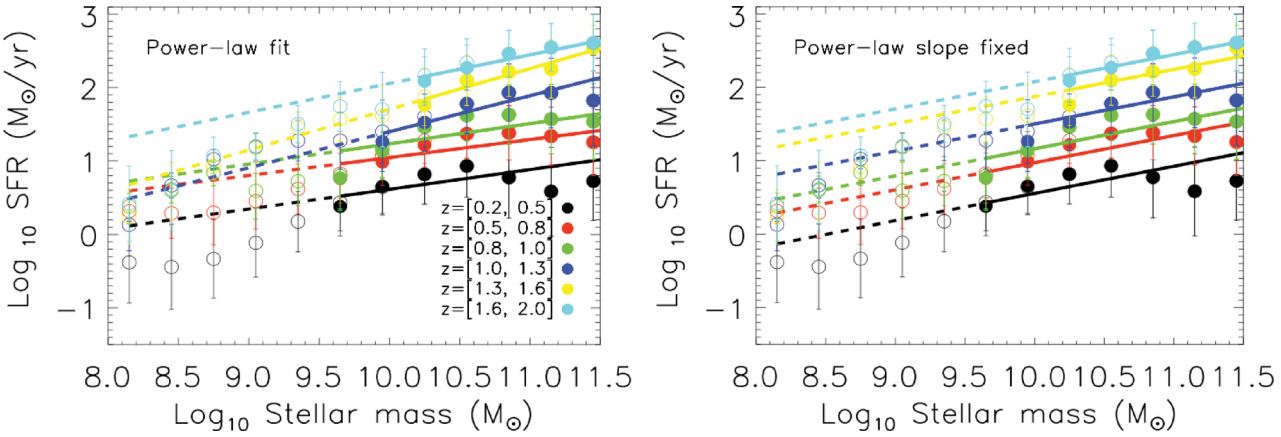


**Figure 2.** The redshift evolution of the average SFR as a function of stellar mass for star-forming galaxies and the best-fitting power law in each redshift slice. The open (filled) circles represent values which are derived from samples below (above) the completeness limit in ECDFS, COSMOS and EGS (see Table 1). Errors include the field-to-field variations and photometric redshift uncertainty. In the left-hand panel, both parameters in the power law are allowed to vary. In the right-hand panel, the power-law slope is fixed to 0.37 which is the average value over different redshift slices. Note that the power-law fitting is only applied to the filled circles. The best-fitting parameters in both panels are listed in Table 2.

probed. At $m_* > 10^8\,\mathrm{M}_\odot$, satellite galaxies make up 18 per cent of the entire population. The fraction of satellite galaxies decreases rapidly with increasing stellar mass at the high-mass end. At $m_* > 5 \times 10^{10}\,\mathrm{M}_\odot$, satellite galaxies account for 8 per cent of the entire population while at $m_* > 10^{11}\,\mathrm{M}_\odot$, the fraction of satellites is <1 per cent. The projected correlation functions in five stellar mass bins (Li et al. 2006) are compared with the best fit from our CSMF in the right-hand panel in Fig. 3. The 1-halo term (due to galaxy pairs residing in the same haloes) dominates the clustering signal on small scales and the 2-halo term (due to galaxy pairs in separate haloes) dominates the clustering signal on large scales. The transition between the 1-halo and 2-halo term is at $\sim 1/h^{-1}$ Mpc$^{-1}$ in all stellar mass bins. The large-scale 2-halo term (proportional to the linear bias factor) increases with $m_*$ indicating more massive galaxies reside in more massive haloes. In the two lowest stellar mass bins, the predicted clustering signal lies below the measured

**Table 1.** Stellar-mass-selected samples in COSMOS, EGS and ECDFS. The columns show redshift range and stellar mass $m_*$ limit in each field above which our samples are regarded as representative.

| $z$ range | COSMOS $\log(m_*[\mathrm{M}_\odot])$ | ECDFS $\log(m_*[\mathrm{M}_\odot])$ | EGS $\log(m_*[\mathrm{M}_\odot])$ |
|---|---|---|---|
| $z = [0.2, 0.5]$ | 9.8 | 9.5 | 9.9 |
| $z = [0.5, 0.8]$ | 10.1 | 9.5 | 10.0 |
| $z = [0.8, 1.0]$ | 10.1 | 9.5 | 10.1 |
| $z = [1.0, 1.3]$ | 10.2 | 9.7 | 10.1 |
| $z = [1.3, 1.6]$ | 10.4 | 9.9 | 10.3 |
| $z = [1.6, 2.0]$ | 10.7 | 10.1 | 10.6 |

on the smallest scales ($\lesssim 0.2\,h^{-1}$ Mpc). It cannot be due to our particular choice of the galaxy density profile inside a dark matter halo because we do not see the same effect in the three more massive mass bins. A full investigation of the cause is deferred until the full covariance matrix of the correlation function is available.

With the parameters in the CSMF tuned by the galaxy abundance and clustering data, we can now predict the average total stellar mass as a function of halo mass which can be calculated from the best-fitting CSMF,

$$\begin{aligned}\langle m_*\rangle_{\rm total} &= \int m_* \times \Phi(m_*|M_{\rm h})\mathrm{d}m_* \\ &= \int m_* \times [\Phi_{\rm cen}(m_*|M_{\rm h}) + \Phi_{\rm sat}(m_*|M_{\rm h})]\mathrm{d}m_* \\ &= \langle m_*\rangle_{\rm cen} + \langle m_*\rangle_{\rm sat}, \end{aligned} \quad (2)$$

which is plotted in Fig. 4. The average stellar mass of the central galaxies as a function of halo mass from our CSMF model agrees reasonably well with constraints from galaxy–galaxy lensing (Mandelbaum et al. 2006; Hoekstra 2007), satellite dynamics (Conroy et al. 2007; More et al. 2011) and galaxy group catalogues (Yang et al. 2009). Our result on the stellar-to-halo mass relation for central galaxies also agrees well with other empirical models, i.e. Moster et al. (2010) and Behroozi et al. (2010).[3] Both Moster et al. (2010) and Behroozi et al. (2010) fit to the observed SMF only. The good agreement between different empirical models indicates that an accurate SMF is the most important constraint in determining the statistical relation between $m_*$ and $M_{\rm h}$. In our CSMF model, the average $m_*$ of the central galaxies grows roughly as $M_{\rm h}^{1.16}$ at the low-mass end and as $M_{\rm h}^{0.71}$ at the high-mass end. The characteristic halo mass for central galaxies in our model, which is where the low- and high-mass power laws meet, is $5 \times 10^{11}\,\mathrm{M}_\odot$. The corresponding stellar mass at the characteristic halo mass is $\sim 10^{10}\,\mathrm{M}_\odot$, which is where local galaxies are found to divide into two distinct families with less massive galaxies showing younger stellar populations, optically blue colours and disc-like morphologies, and more massive galaxies exhibiting older stellar populations, optically red colours and more bulge-like morphology (Kauffmann et al. 2004). Therefore, the different stellar mass build-up history, indicated by the different $m_*-M_{\rm h}$ relation below and above $M_{\rm h} = 5 \times 10^{11}\,\mathrm{M}_\odot$, may explain the observed division in galaxy properties below and above $m_* \sim 10^{10}\,\mathrm{M}_\odot$.

[3] Moster et al. (2010) use an almost identical CSMF formalism to what is used in this paper. Essentially, it is a double power law connected at some characteristic mass scale. Behroozi et al. (2010) use a different CSMF formalization. The main difference is that for high-mass galaxies, the $m_*-M_{\rm h}$ relation asymptotes to a sub-exponential function instead of a power law.

### 3.2 The stellar-to-halo mass relation at high $z$

In Table 3, we list a series of volume- and stellar-mass-limited subsamples in six redshift bins in COSMOS and EGS. The projected correlation function for each subsample in COSMOS and EGS is plotted inFig. 5. More details on how projected correlation functions are calculated can be found in Appendix B. In redshift bins where multiple stellar-mass-limited subsamples exist, it seems that more massive galaxies generally show stronger clustering although the large errors prevent any firm conclusions to be drawn. This is consistent with Meneux et al. (2009) who studied the clustering dependence on $m_*$ in the redshift bin $z = [0.2, 1]$ using the first 10K redshifts from the $z$COSMOS survey and found a mild dependence on $m_*$ especially on small scales (see Fig. 6).

We derive the $m_*-M_{\rm h}$ relation for the local Universe by fitting to both the spatial density and clustering of galaxies. At high $z$, however, we will only use the SMFs and not the correlation functions presented above. This is because the correlation function is extremely sensitive to cosmic variance. A large difference in the correlation functions between COSMOS and VIRMOS VLT Deep Survey (VVDS) was reported in Meneux et al. (2009). Also, the flat shape in the measured zCOSMOS correlation functions (shown in Fig. 6) over the redshift range $z = [0.6, 1.0]$ has been attributed to an overabundance of high-density regions (de la Torre et al. 2010). We show the measured SMFs in Pérez-González et al. (2008) based on a combined sample of 3.6 and 4.5 μm selected sources in the HDF-N, the CDF-S and the Lockman Hole and the best-fitting from our CSMF model in Fig. 7. Note that in comparing to the observed SMF, the predicted SMF (i.e. the intrinsic SMF) from the best-fitting CSMF model (the black dashed line in Fig. 7) has been convolved with a log-normal distribution with its width set to 0.3 dex (Pérez-González et al. 2008) to account for statistical errors in the observational estimate of stellar mass. The SMF increases over time but mostly in low-mass systems. The contribution from satellites also grows over time. In Fig. 5, the projected correlation functions in COSMOS and EGS are compared with the predicted correlation functions from our best-fitting CSMF. There is relatively good overall agreement between the two. On large scales, the measured correlation function falls under the predicted curve, which is due to integral constraint. If the galaxy number density fluctuations in the probed volume are smaller than the average over a cosmologically representative volume, then the measured correlation function will be biased low by a constant, which is equal to the fractional variance of the number counts in cells. This effect is significant if the survey field is small. In Fig. 6, we compare the projected correlation functions of stellar-mass-limited samples from the $z$COSMOS 10K sample (Meneux et al. 2009) with the predicted clustering from our model and again find relatively good overall agreement.

We plot the average $m_*-M_{\rm h}$ relation as a function of $M_{\rm h}$ for central galaxies in the left-hand panel in Fig. 8. The characteristic halo mass scale in the $m_*-M_{\rm h}$ relation for central galaxies has increased with increasing redshift, changing from $\sim 5.0 \times 10^{11}\,\mathrm{M}_\odot$ at $z = 0.1$ to $\sim 1.1 \times 10^{12}\,\mathrm{M}_\odot$ at $z = 1.8$. In the right-hand panel of Fig. 8, we plot the $m_*$-to-$M_{\rm h}$ (a measure of the integrated star-formation efficiency) as a function of $M_{\rm h}$ for central galaxies. It is clear that the integrated star-formation efficiency is low in both low-mass and high-mass haloes in all redshift slices at $0 < z < 2$. In low-mass haloes, star-formation efficiency is suppressed possibly due to supernova feedback which can reheat the interstellar stellar medium, heat gas in the dark matter halo or even eject gas altogether (Springel & Hernquist 2003; Brooks et al. 2007; Ceverino & Klypin 2009). In high-mass haloes, star-formation efficiency is

**Table 2.** A two-parameter fit of the form $\psi(\mathrm{M}_\odot\ \mathrm{yr}^{-1}) = \alpha \times (m_*/\mathrm{M}_\odot)^\beta$ to the stellar mass dependence of the average SFR for star-forming galaxies, averaged over ECDFS, COSMOS and EGS.

| Power-law fit | | | | Power-law slope fixed | | |
|---|---|---|---|---|---|---|
| $\Delta z$ | $\log_{10}\alpha\ (\mathrm{yr}^{-1})$ | $\beta$ | $\chi^2$/d.o.f. | $\log_{10}\alpha\ (\mathrm{yr}^{-1})$ | $\beta$ | $\chi^2$/d.o.f. |
| $z = [0.2, 0.5]$ | $-2.08 \pm 2.26$ | $0.27 \pm 0.21$ | 0.17 | $-3.14 \pm 0.46$ | 0.37 | 0.20 |
| $z = [0.5, 0.8]$ | $-1.41 \pm 1.92$ | $0.25 \pm 0.18$ | 0.18 | $-2.73 \pm 0.37$ | 0.37 | 0.26 |
| $z = [0.8, 1.0]$ | $-1.52 \pm 2.04$ | $0.28 \pm 0.19$ | 0.30 | $-2.54 \pm 0.37$ | 0.37 | 0.34 |
| $z = [1.0, 1.3]$ | $-3.52 \pm 2.66$ | $0.49 \pm 0.25$ | 0.09 | $-2.20 \pm 0.40$ | 0.37 | 0.14 |
| $z = [1.3, 1.6]$ | $-3.85 \pm 3.47$ | $0.56 \pm 0.32$ | 0.05 | $-1.83 \pm 0.36$ | 0.37 | 0.14 |
| $z = [1.6, 2.0]$ | $-1.86 \pm 2.65$ | $0.39 \pm 0.24$ | 0.03 | $-1.62 \pm 0.35$ | 0.37 | 0.03 |

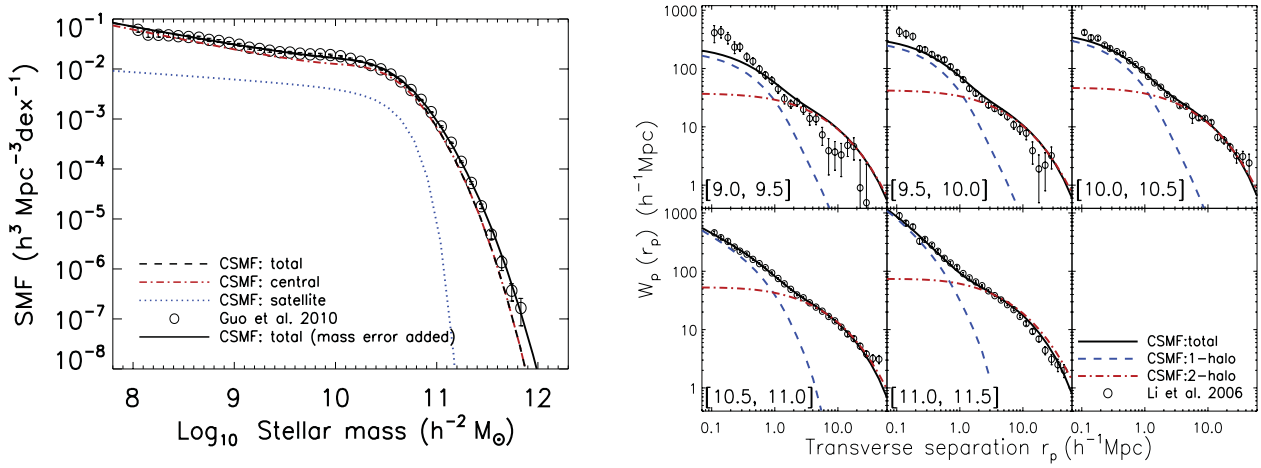


**Figure 3.** Left: the measured SMF of the local Universe based on SDSS DR7 (Guo et al. 2010) compared with the SMF derived from our best-fitting CSMF. Central galaxies (the red dot–dashed line) dominate the SMF over the entire mass range probed. Right: the measured projected correlation functions of the SDSS galaxies in different stellar mass bins (Li et al. 2006) compared with the correlation functions derived from our best-fitting CSMF. Note that the stellar mass bins shown in each panel are calculated with $h = 0.7$. For example, the top-left panel shows the projected correlation function of galaxies in the stellar mass bin $\log_{10} m_*(\mathrm{M}_\odot) = [9.0, 9.5]$ assuming $h = 0.7$.

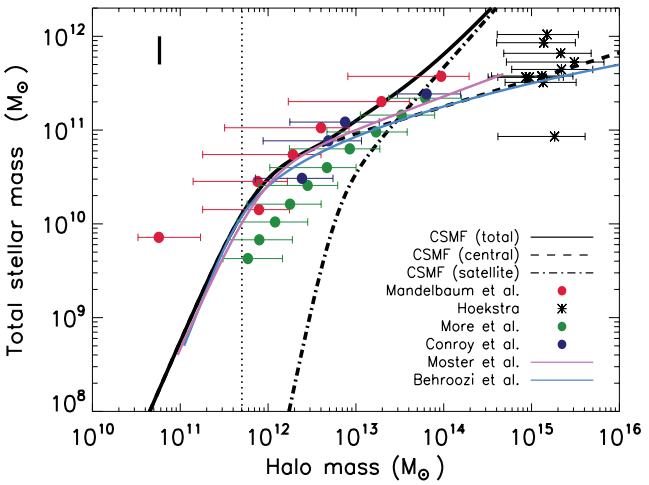


**Figure 4.** The predicted average total stellar mass as a function of $M_\mathrm{h}$ derived from the best-fitting CSMF of the local Universe. The predicted average stellar mass of central galaxies agrees reasonably well with results from galaxy–galaxy lensing (Mandelbaum et al. 2006; Hoekstra 2007), satellite kinematics (Conroy et al. 2007; More et al. 2011) and other empirical models of the stellar-to-halo mass relation (Behroozi et al. 2010; Moster et al. 2010). The vertical bar on the left indicates the typical error in $m_*$. The vertical dotted line marks the characteristic $M_\mathrm{h}$ in the $m_*-M_\mathrm{h}$ relation for central galaxies.

**Table 3.** Volume-limited and stellar-mass-selected subsamples in COSMOS and EGS used to calculate correlation functions. The columns show sample name, redshift range, number of galaxies and stellar mass range. Note that the number of galaxies in each sample varies slightly in different Monte Carlo realizations.

| Sample | $z$ range | $N_\mathrm{gal}$ | $\log_{10} m_*\ (\mathrm{M}_\odot)$ |
|---|---|---|---|
| z1M1 (COSMOS) | $z = [0.2, 0.5]$ | 2117 | [9.8, 10.1] |
| z1M2 (COSMOS) | $z = [0.2, 0.5]$ | 2025 | [10.1,10.4] |
| z1M3 (COSMOS) | $z = [0.2, 0.5]$ | 2175 | >10.4 |
| z2M1 (COSMOS) | $z = [0.5, 0.8]$ | 2311 | [10.1, 10.4] |
| z2M2 (COSMOS) | $z = [0.5, 0.8]$ | 2641 | [10.4, 10.6] |
| z2M3 (COSMOS) | $z = [0.5, 0.8]$ | 2369 | >10.6 |
| z3M1 (COSMOS) | $z = [0.8, 1.0]$ | 4821 | [10.1, 10.5] |
| z3M2 (COSMOS) | $z = [0.8, 1.0]$ | 5051 | >10.5 |
| z4M1 (COSMOS) | $z = [1.0, 1.3]$ | 4111 | [10.2, 10.6] |
| z4M2 (COSMOS) | $z = [1.0, 1.3]$ | 3950 | >10.6 |
| z5M1 (COSMOS) | $z = [1.3, 1.6]$ | 3867 | >10.4 |
| z6M1 (COSMOS) | $z = [1.6, 2.0]$ | 2425 | >10.7 |
| z1M1 (EGS) | $z = [0.2, 0.5]$ | 2064 | >9.9 |
| z2M1 (EGS) | $z = [0.5, 0.8]$ | 2186 | [10.0, 10.2] |
| z2M2 (EGS) | $z = [0.5, 0.8]$ | 2045 | >10.2 |
| z3M1 (EGS) | $z = [0.8, 1.0]$ | 2650 | >10.1 |
| z4M1 (EGS) | $z = [1.0, 1.3]$ | 2965 | >10.1 |
| z5M1 (EGS) | $z = [1.3, 1.6]$ | 2333 | >10.3 |
| z6M1 (EGS) | $z = [1.6, 2.0]$ | 1731 | >10.6 |

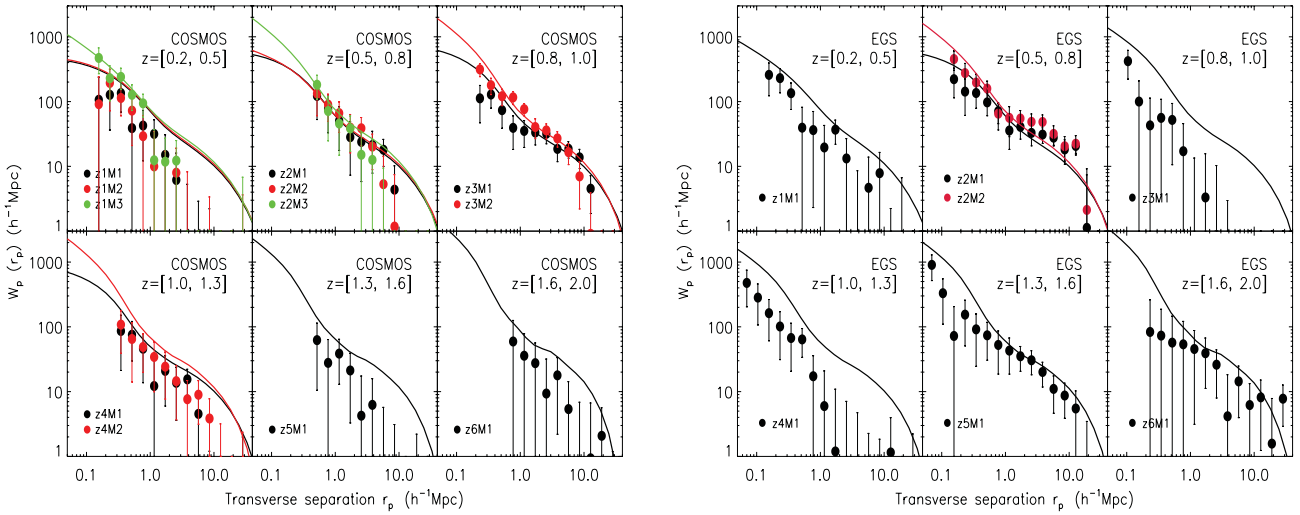


**Figure 5.** The projected correlation functions of stellar-mass-limited subamples listed in Table 3 in six redshift bins $z1 = [0.2, 0.5]$, $z2 = [0.5, 0.8]$, $z3 = [0.8, 1.0]$, $z4 = [1.0, 1.3]$, $z5 = [1.3, 1.6]$ and $z6 = [1.6, 2.0]$. The solid lines show the predicted correlation function from our best-fitting CSMF. Error bars include both the bootstrapping error and the photometric redshift error. In redshift bins where multiple stellar-mass-limited subsamples exist, more massive galaxies seem to show a higher clustering amplitude than less massive galaxies.

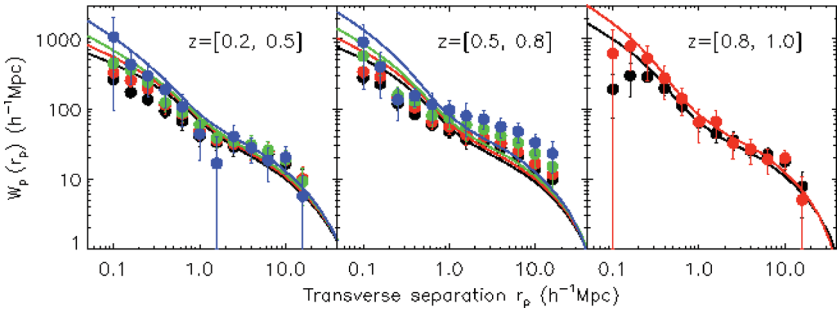


**Figure 6.** The projected correlation functions of stellar-mass-limited samples (black points: galaxies with $\log(m_*/M_\odot) \geq 9.0$; red points: $\log(m_*/M_\odot) \geq 9.5$; green points: $\log(m_*/M_\odot) \geq 10.0$; blue points: $\log(m_*/M_\odot) \geq 10.5$) in three redshift bins $z1 = [0.2, 0.5]$, $z2 = [0.5, 0.8]$ and $z3 = [0.8, 1.0]$ from the $z$COSMOS 10K sample (Meneux et al. 2009). The solid lines denote the predicted correlation function from our best-fitting CSMF. The flat shape in the measured $z$COSMOS correlation functions in the middle panel has been explained by an overabundance of high-density regions (de la Torre et al. 2010).

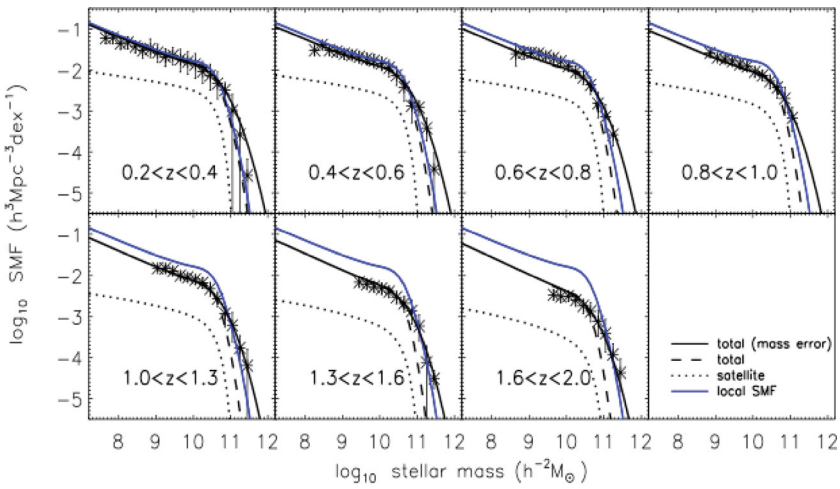


**Figure 7.** The measured SMFs in different redshift bins from Pérez-González et al. (2008). The redshift range is indicated in each panel. The dashed black line in each panel shows the underlying SMF predicted from the best-fitting CSMF. The solid black line in each panel denotes the convolution of the dashed black line with a log-normal distribution which represents the statistical error in the observational estimate of stellar mass. The blue line shows the present-day SMF (i.e. the black dashed line in the left-hand panel of Fig. 3).

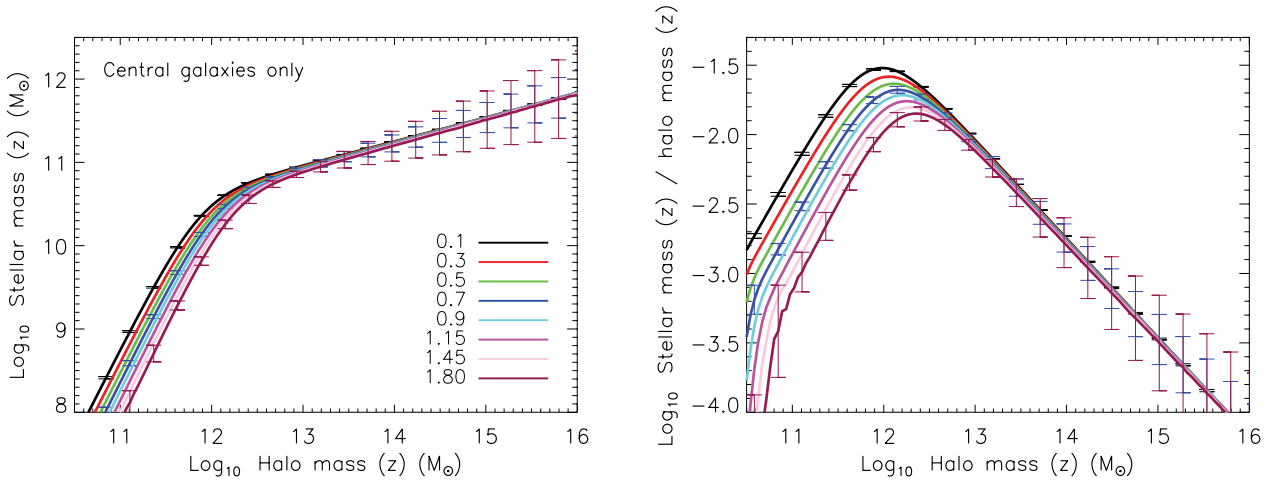


**Figure 8.** Left: the predicted average stellar mass of central galaxies as a function of halo mass from the best-fitting CSMF. For clarity, we only plot errors on a few selected redshift slices and halo mass bins. Different lines are colour-coded by redshift as indicated in the panel. The characteristic halo mass scale for central galaxies increases with increasing redshift. Right: the average stellar-to-halo mass ratio for the central galaxies versus the host halo mass predicted from the best-fitting CSMF. It is clear that the star-formation efficiency is low in both low-mass and high-mass haloes and the peak in the stellar-to-halo mass ratio shifts to lower mass haloes over time.

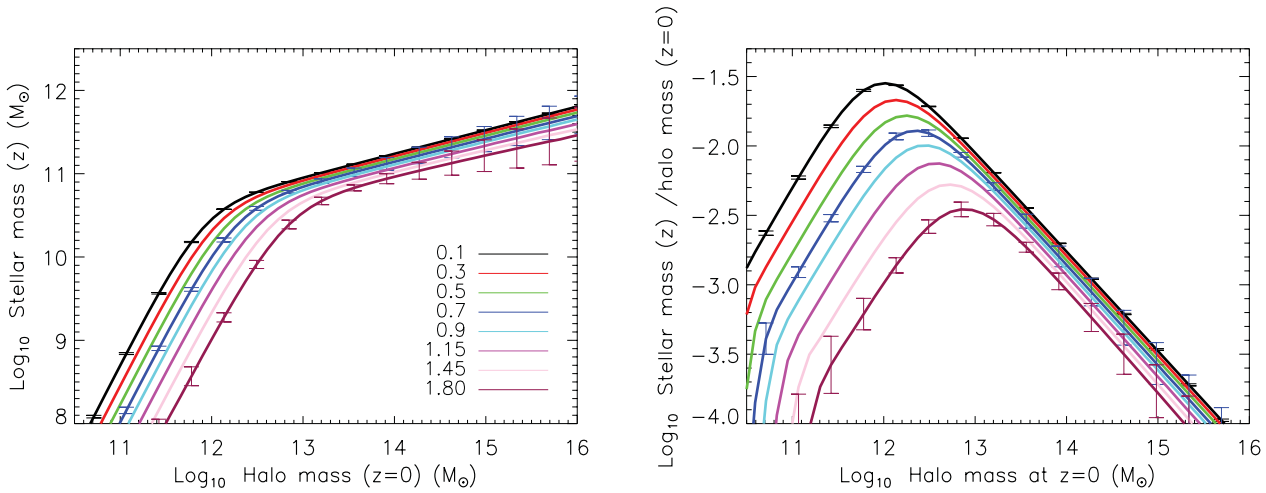


**Figure 9.** Left: the predicted average stellar mass of the central galaxies as a function of halo mass at the present day. The halo mass is evolved to $z = 0$ using the halo mass accretion history from Fakhouri et al. (2010). The evolution of the stellar content as a function of halo mass is shown. Different lines are colour-coded by redshift as indicated in the panel. Right: the average stellar-to-halo mass ratio versus halo mass at the present day. The build-up of stellar mass happened early on in massive haloes.

also suppressed possibly due to gravitational heating (Dekel & Birnboim 2006 ; Khochfar & Ostriker 2008) and/or feedback from AGN which transfers energy to the halo gas (Bower et al. 2006; Croton et al. 2006; Monaco et al. 2007). The peak of the average stellar-to-halo mass ratio for central galaxies has shifted towards lower mass haloes over time.

In Fig. 9, we plot the stellar mass build-up history as a function of halo mass at the present day by evolving $M_{\rm h}$ at a particular redshift to $M_{\rm h}$ at $z = 0$ using the halo mass accretion rate from Fakhouri et al. (2010),

$$\left\langle \frac{{\rm d}M_{\rm h}}{{\rm d}t} \right\rangle = 46.1 \left( \frac{M_{\rm h}}{10^{12}} \right)^{1.1} (1 + 1.11z)\sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}. \quad (3)$$

So we can trace the evolution of the stellar content in the same halo along any vertical line in Fig. 9. It is clear that the stellar mass assembly happened much earlier in massive haloes than in less massive haloes. In haloes more massive than $10^{13}\,{\rm M}_\odot$ (the present-day value), the stellar mass of the central galaxies has increased by at most a factor of a few. But in less massive haloes, the stellar mass of the central galaxies has grown by an order of magnitude or more. This is consistent with the downsizing scenario of galaxy formation.

## 4 EHM: 2. CONNECTING STELLAR MASS, SFR AND HALO MASS

Now we can extend the CSMF to the 2D distribution $\Phi(\psi, m_*|M_{\rm h})$, which specifies the number of galaxies as a function of $m_*$ and $\psi$ at fixed $M_{\rm h}$. Using conditional probability theory, one can show that

$$\Phi(\psi, m_*|M_{\rm h}) = \Phi(m_*|M_{\rm h}) \times \Phi(\psi|m_*, M_{\rm h}). \quad (4)$$

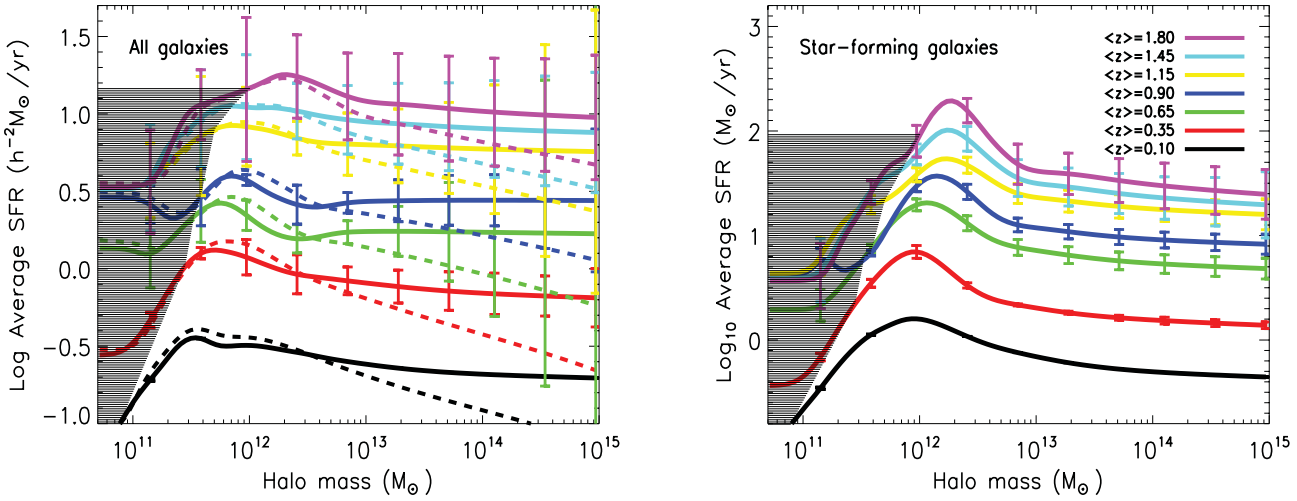


**Figure 10.** Left: the average SFR as a function of $M_{\rm h}$. Error bars include the uncertainty in the parametrized stellar-to-halo mass relation, the field-to-field variation in the SFR distribution as a function of stellar mass and the photometric redshift error. At $z = 0.1$, the errors on the SFR–$M_{\rm h}$ relation are very small because the error bars only include the uncertainty in the parametrized stellar-to-halo mass relation. Different lines are colour-coded by redshift as indicated. The solid/dashed lines correspond to the $\psi - M_{\rm h}$ relation derived from Scenario A/B. The hatched regions indicate the $M_{\rm h}$ range where we are not able to derive reliable constraints on the $\psi - M_{\rm h}$ relation due to the increasingly limited $m_*$ range probed towards higher $z$. Right: similar to the left-hand panel but for star-forming galaxies only. Since the SFR distribution at a given stellar mass for star-forming galaxies is independent of environment, only Scenario A is plotted.

If the distribution of SFR is only dependent on $m_*$ and at most weakly dependent on $M_{\rm h}$, then one can assume

$$\Phi(\psi, m_*|M_{\rm h}) \approx \Phi(m_*|M_{\rm h}) \times \Phi(\psi|m_*). \quad (5)$$

We will refer to this simplification as Scenario A.

However, it is important to realize that the distribution of SFR at fixed $m_*$ may be different in haloes of different masses, which is a measure of the Mpc-scale environment. Using group catalogues constructed from the SDSS DR5 (Yang et al. 2007), Kimm et al. (2009) studied the fraction of passive galaxies, $f_{\rm passive}$, as a function of $m_*$ and $M_{\rm h}$. Within the error bars, it is difficult to tell whether $f_{\rm passive}$ at fixed $m_*$ has any significant dependence on $M_{\rm h}$. However, Peng et al. (2010) using both the SDSS and $z$COSMOS data sets found that the SFR of star-forming galaxies at fixed $m_*$ is completely independent of environment (measured by the fifth nearest neighbour density estimator),[4] but $f_{\rm passive}$ depends on environment even at fixed $m_*$. Therefore, in this paper, we adopt a second scenario in building the 2D distribution in the $(\psi, m_*)$ plane as a function of halo mass. We assume that the fraction of passive galaxies at fixed $m_*$ has a power dependence on $M_{\rm h}$, i.e. $f_{\rm passive}(M_{\rm h}|m_*) \propto M_{\rm h}^{\eta(m_*)}$. Furthermore, we assume that all galaxies are passive in very massive haloes (corresponding to the most massive rich clusters), i.e. $f_{\rm passive} = 1$ at $M_{\rm h} = 10^{15}\,{\rm M}_\odot$. Since we know the overall $f_{\rm passive}$ in a given stellar mass bin, we can work out the power-law dependence $\eta(m_*)$. Under this assumption, the SFR distribution at fixed $m_*$ and $M_{\rm h}$ can be derived from the SFR distribution at fixed $m_*$ but with $f_{\rm passive}$ modulated by halo mass, i.e.

$$\Phi(\psi|m_*, M_{\rm h}) \approx \Phi(\psi|m_*) f_{\rm passive}(M_{\rm h}|m_*). \quad (6)$$

We will refer to this simplification as Scenario B.

[4] Since the SFR distribution at a given stellar mass is independent of environment for star-forming galaxies, we will only need to use Scenario A to connect SFR with halo mass for the star-forming galaxy population.

In Fig. 10, we plot the average total SFR as a function of $M_{\rm h}$ at various redshifts. The left-hand panel is for all galaxies and the right-hand panel is for star-forming galaxies only (as defined in Section 2.4). Error bars include the uncertainty in the parametrized stellar-to-halo mass relation, the field-to-field variation in the SFR distribution as a function of stellar mass and the photometric redshift error. At $z = 0.1$, the errors on the SFR–$M_{\rm h}$ relation are very small because the error bars only include the uncertainty in the parametrized stellar-to-halo mass relation. The average SFR is higher/lower in less/more massive haloes in Scenario B than in Scenario A. This is because Scenario B assumes that the $f_{\rm passive}$ increases with increasing $M_{\rm h}$. However, the difference in the SFR as a function of $M_{\rm h}$ between the two scenarios is small and does not affect the qualitative conclusions drawn in this paper. We can see that the intensity of star-forming activity in haloes in the probed mass range has steadily decreased as a function of time, dropping by over one order of magnitude from $z \sim 2$ to $z \sim 0$. The peak in SFR shifts from $M_{\rm h}$ just over $10^{12}\,{\rm M}_\odot$ at $z \sim 2$ to just below $10^{12}\,{\rm M}_\odot$ at $z \sim 0.1$, in qualitative agreement with Fig. 8 where the peak of the stellar-to-halo mass ratio (a measure of the integrated star-formation efficiency) is shown to shift towards lower mass haloes over time. At a given redshift, haloes in the mass range between a few times $10^{11}\,{\rm M}_\odot$ and a few times $10^{12}\,{\rm M}_\odot$ are the most efficient at hosting star formation. Again, this is consistent with Fig. 8 which shows that the integrated star-formation efficiency is low in both low- and high-mass haloes and peaks at $\sim 10^{12}\,{\rm M}_\odot$.

In Fig. 11, $M_{\rm h}$ is evolved to $z = 0$ using the halo mass accretion history derived from numerical simulations (Fakhouri et al. 2010), i.e. equation (6). So we can trace the star-formation history in the same halo along any vertical line. One can read off the evolutionary sequence of different populations of galaxies. Galaxies that are forming stars most actively at $z \sim 2$ have evolved into populations that reside in group-like environments at the present day and galaxies that are forming stars most actively at the present day generally reside in field environment. This explains the reversal of the SFR–density relation at high redshift first presented in Elbaz

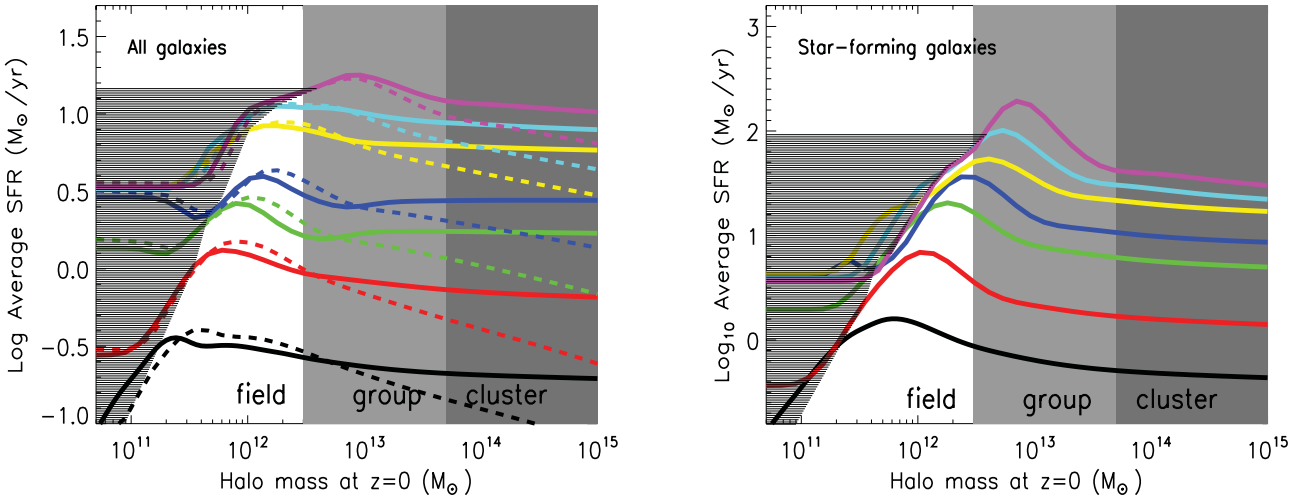


**Figure 11.** Left: the average SFR as a function of halo mass at the present day. $M_{\rm h}$ is evolved to $z = 0$ using the halo accretion history from numerical simulations (Fakhouri et al. 2010). The solid/dashed lines correspond to the $\psi - M_{\rm h}$ relation derived from Scenario A/B. Along any vertical line, we can trace the evolution of the SFR in the same halo. The dark grey/light grey/white region indicates $M_{\rm h}$ range typically associated with cluster/group/field environment. For clarity, error bars are not shown here. Right: similar to the left-hand panel but for star-forming galaxies only. It is clear that the most actively star-forming galaxies at $z \sim 2$ reside in group-like environment and they evolve into quiescent galaxies in groups at the present day.

et al. (2007) and strongly supports previous claims that the most powerful starbursts at $z \sim 2$ (i.e. sub-mm galaxies) have evolved into today's elliptical galaxies in dense environment (e.g. Lilly et al. 1999; Smail et al. 2004; Swinbank et al. 2006). It is worth pointing out that our results on the redshift evolution of the average SFR as a function of halo mass are in good qualitative agreement with some recent results in the literature (Behroozi, Wechsler & Conroy 2012; Moster, Naab & White 2013). A detailed quantitative comparison (e.g. the impact of different methodology, different observations used to constrain the empirical model, etc.) is beyond the scope of this paper.

## 5 DISCUSSIONS AND CONCLUSIONS

In the last 10 years there has been an explosion of spectroscopic and multiwavelength photometric data charting the star-formation history and stellar mass build-up over a large fraction of cosmic time. And now the advent of *Herschel* allows us to reliably probe the obscured star-formation activity in large numbers of high-$z$ galaxies. In the near future, powerful space- and ground-based facilities will dramatically increase sample size and allow robust measurements of galaxy properties to be made at even higher redshift.

In this paper, we present an EHM of galaxy evolution which links stellar mass ($m_*$) and SFR of galaxies to their underlying host halo mass ($M_{\rm h}$) from the local Universe to $z \sim 2$. While the empirical relation between $m_*$ and $M_{\rm h}$ has been constructed based on observations before, this is the first time the relation between $\psi$ and $M_{\rm h}$ has been constructed from observational data over 80 per cent of cosmic time. The *Herschel*-SPIRE observations obtained as part of the HerMES are crucial for obtaining accurate SFR estimates for dusty star-forming galaxies at high $z$.

The EHM is built through two steps.

(i) First, we build the CSMF $\Phi(m_*|M_{\rm h})$, which specifies the average number of galaxies as a function of $m_*$ in a halo of a given mass. The CSMF, by construction, fits the SMF and the projected correlation functions as a function of stellar mass in the local Universe and the SMFs in various redshift slices in the distant Universe. The predicted clustering properties from our best-fitting CSMF as a function of redshift also agree reasonably well with the measured correlation functions at high $z$ (modulo integral constraint and cosmic variance effect).

(ii) Secondly, we extend the CSMF to the joint distribution in $\psi$ and $m_*$ as a function of halo mass, $\Phi(\psi, m_*|M_{\rm h})$, by incorporating the distribution of SFR at fixed $m_*$. We have used two scenarios in building $\Phi(\psi, m_*|M_{\rm h})$. Scenario A assumes that $m_*$ plays the most important role in determining the SFR distribution of galaxies and the effect of $M_{\rm h}$ at fixed $m_*$ is negligible. Scenario B assumes that the SFR distribution at fixed $m_*$ has a power-law dependence on $M_{\rm h}$. The difference in the resulting $\psi - M_{\rm h}$ relation is small between the two different scenarios and does not affect the main conclusions presented in the paper.

Combining the halo accretion history from numerical simulations and the 2D distribution of $m_*$ and $\psi$ as a function of $M_{\rm h}$ in various redshift slices $\Phi(\psi, m_*|M_{\rm h}, z)$, we can trace the stellar mass growth and the evolution of SFR in different haloes. Our most important findings are as follows.

(i) The intensity of the star-forming activity in haloes in the probed mass range has steadily decreased over time, dropping by over one order of magnitude from $z \sim 2$ to $z \sim 0$.

(ii) At each redshift, haloes in the mass range between a few times $10^{11}\,{\rm M}_\odot$ and a few times $10^{12}\,{\rm M}_\odot$ are the most efficient at hosting star formation, consistent with the optimum halo mass scale for star formation predicted from numerical simulations.

(iii) The peak of SFR and the peak of the stellar-to-halo mass ratio (a measure of the integrated star-formation efficiency) shift to lower mass haloes as redshift decreases.

(iv) Galaxies that are forming stars most actively at $z \sim 2$ have evolved into quiescent galaxies in group-like environments at the present day.

To further constrain the physical processes responsible for the $\psi - M_{\rm h}$ relation and its evolution with redshift, future work is needed to investigate the role of three main suspects: molecular gas content and evolution, feedback from central massive black holes and

environmental effects on star formation. The advent of the Atacama Large Millimeter/submillimeter Array, the Expanded Very Large Array (Perley et al. 2011) and the Northern Extended Millimeter Array means that we are now in a position to be able to measure the evolution of the molecular gas content in a statistically significant sample of galaxies with moderate SFRs. The feedback from growing black holes may also impact the star-formation activity in massive haloes ($M_{\rm h} > 10^{13}\,{\rm M}_{\odot}$), as required in order to reproduce the observed stellar mass and luminosity functions of galaxies in numerical simulations and semi-analytic models (e.g. Bower et al. 2006, 2008; Croton et al. 2006). We will extend the EHM to include the empirical relation between AGNs and halo mass in a future paper to statistically investigate star formation–black hole co-evolution. Finally, the impact of environment can be studied through galaxy group and cluster catalogues over a large redshift range and will be presented in a separate paper.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for constructive comments. We also thank Cheng Li for providing the SDSS clustering measurements and the stellar mass function, Rachel Mandelbaum for the stellar mass $m_*$–halo mass $M_{\rm h}$ relation from weak lensing, Surhud More for the $m_*-M_{\rm h}$ relation from satellite kinematics, Alexi Leauthaud for the $m_*-M_{\rm h}$ relation, Samir Salim for the conditional dependence of specific star-formation rate on $m_*$ in the local Universe, Benjamin Moster for useful discussions on the CSMF, Anthony Lewis for useful discussion on MCMC methods and Peder Norberg for general discussions.

LW acknowledges support from UK's Science and Technology Facilities Council grant ST/F002858/1 and an ERC StG grant (DEGAS-259586). SJO is supported by UK's Science and Technology Facilities Council grant ST/F002858/1. The data presented in this paper will be released through the Herschel data base in Marseille HeDaM (hedam.oamp.fr/herMES). SPIRE has been developed by a consortium of institutes led by Cardiff Univ. (UK) and including Univ. Lethbridge (Canada); NAOC (China); CEA, LAM (France); IFSI, Univ. Padua (Italy); IAC (Spain); Stockholm Observatory (Sweden); Imperial College London, RAL, UCL-MSSL, UKATC, Univ. Sussex (UK); and Caltech, JPL, NHSC, Univ. Colorado (USA). This development has been supported by national funding agencies: CSA (Canada); NAOC (China); CEA, CNES, CNRS (France); ASI (Italy); MCINN (Spain); SNSB (Sweden); STFC (UK); and NASA (USA).

## REFERENCES

Arnouts S. et al., 2002, MNRAS, 329, 355
Arnouts S. et al., 2007, A&A, 476, 137
Barro G. et al., 2011a, ApJS, 193, 13
Barro G. et al., 2011b, ApJS, 193, 30
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2012, arXiv:1207.6105
Bell E. F. et al., 2005, ApJ, 625, 23
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nat, 311, 517
Bouwens R. J. et al., 2010, ApJ, 709, L133
Bower R. G. et al., 2006, MNRAS, 370, 645
Bower R. G., McCarthy I. G., Benson A. J., 2008, MNRAS, 390, 1399
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brooks A. M., Governato F., Booth C. M., Willman B., Gardner J. P., Wadsley J., Stinson G., Quinn T., 2007, ApJ, 655, L17
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D. et al., 2000, ApJ, 533, 682
Cardamone C. N. et al., 2010, ApJS, 189, 270
Ceverino D., Klypin A., 2009, ApJ, 695, 292
Chabrier G., 2003, PASP, 115, 763
Chary R., Elbaz D., 2001, ApJ, 556, 562
Conroy C. et al., 2007, ApJ, 654, 153
Croton D. J. et al., 2006, MNRAS, 365, 11
Daddi E. et al., 2007, ApJ, 670, 156
Damen M. et al., 2011, ApJ, 727, 1
de la Torre S. et al., 2010, MNRAS, 409, 867
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dickinson M., Papovich C., Ferguson H. C., Budavári T., 2003, ApJ, 587, 25
Elbaz D. et al., 2007, A&A, 468, 33
Elbaz D. et al., 2010, A&A, 518, L29
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189
Griffin M. J. et al., 2010, A&A, 518, L3
Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111
Hoekstra H., 2007, MNRAS, 379, 317
Hoekstra H., Yee H. K. C., Gladders M. D., 2004, ApJ, 606, 67
Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142
Ilbert O. et al., 2006, A&A, 457, 841
Ilbert O. et al., 2009, ApJ, 690, 1236
Ilbert O. et al., 2010, ApJ, 709, 644
Karim A. et al., 2011, ApJ, 730, 61
Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713
Kennicutt R. C., Jr, 1998, ARA&A, 36, 189
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Khochfar S., Ostriker J. P., 2008, ApJ, 680, 54
Kimm T. et al., 2009, MNRAS, 394, 1131
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, MNRAS, 368, 21
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Lilly S. J., Le Fevre O., Hammer F., Crampton D., 1996, ApJ, 460, L1
Lilly S. J. et al., 1999, ApJ, 518, 641
Lin Y.-T., Mohr J. J., 2004, ApJ, 617, 879
Lin Y.-T., Mohr J. J., Stanford S. A., 2003, ApJ, 591, 749
Lotz J. M. et al., 2008, ApJ, 672, 177
Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
McKay T. A. et al., 2001, preprint (arXiv:astro-ph/0108013)
McKay T. A. et al., 2002, ApJ, 571, L85
Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715
Meneux B. et al., 2009, A&A, 505, 463
Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nat, 379, 613
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Murphy E. J. et al., 2009, ApJ, 698, 1380
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Noeske K. G. et al., 2007, ApJ, 660, L43
Nordon R. et al., 2010, A&A, 518, L24
Nordon R. et al., 2012, ApJ, 745, 182
Oliver S. J. et al., 2012, MNRAS, 424, 1614
Papovich C. et al., 2007, ApJ, 668, 45
Pascale E. et al., 2009, ApJ, 707, 1740
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peng Y.-j. et al., 2010, ApJ, 721, 193
Perez-Gonzalez P. G. et al., 2008, ApJ, 675, 234
Perley R. A., Chandler C. J., Butler B. J., Wrobel J. M., 2011, ApJ, 739, L1
Pilbratt G. L. et al., 2010, A&A, 518, L1
Rodighiero G. et al., 2010, A&A, 518, L25
Roseboom I. G. et al., 2010, MNRAS, 409, 48

Roseboom I. G. et al., 2012, MNRAS, 419, 2758
Rudnick G. et al., 2003, ApJ, 599, 847
Salim S. et al., 2007, ApJS, 173, 267
Sanders D. B. et al., 2007, ApJS, 172, 86
Schiminovich D. et al., 2005, ApJ, 619, L47
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Seljak U. et al., 2000, MNRAS, 318, 203
Sheldon E. S. et al., 2004, AJ, 127, 2544
Sheldon E. S. et al., 2009, ApJ, 703, 2217
Smail I., Chapman S. C., Blain A. W., Ivison R. J., 2004, ApJ, 616, 71
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Swinbank A. M. et al., 2006, MNRAS, 371, 465
Tinker J. et al., 2008, ApJ, 688, 709
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van den Bosch F. C., Yang X., Mo H. J., 2003, MNRAS, 340, 771
van den Bosch F. C., Norberg P., Mo H. J., Yang X., 2004, MNRAS, 352, 1302
Vikhlinin A. et al., 2006, ApJ, 640, 691
Whitaker K. E. et al., 2011, ApJ, 735, 86
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, MNRAS, 358, 217
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 695, 900
Zehavi I. et al., 2004, ApJ, 608, 16

## APPENDIX A: THE CONDITIONAL STELLAR MASS FUNCTION

Motivated by studies of galaxy groups (Yang et al. 2005), we can divide the CSMF into that of central and satellite galaxies,

$$\Phi(m_*|M_{\rm h}) = \Phi_{\rm cen}(m_*|M_{\rm h}) + \Phi_{\rm sat}(m_*|M_{\rm h}), \tag{A1}$$

where $\Phi_{\rm cen}(m_*|M_{\rm h})$ and $\Phi_{\rm sat}(m_*|M_{\rm h})$ specify the number of central and satellite galaxies as a function of $m_*$ at fixed $M_{\rm h}$, respectively. A log-normal distribution is used to model the CSMF of central galaxies,

$$\Phi_{\rm cen}(m_*|M_{\rm h}) = \frac{1}{\sqrt{2\pi}\ln 10 m_*\sigma_{\rm c}}\exp\left[-\frac{\log^2(m_*/m_{\rm c})}{2\sigma_{\rm c}^2}\right], \tag{A2}$$

where $m_{\rm c}(M_{\rm h})$ is the mean stellar mass of a central galaxy in a halo of mass $M_{\rm h}$ and $\sigma_{\rm c} = 0.2$ dex is the standard deviation. Following Moster et al. (2010), $m_{\rm c}(M_{\rm h})$ is parametrized as

$$m_{\rm c}(M_{\rm h}) = 2M_{\rm h}\left(\frac{m_{\rm c}}{M}\right)_0\left[\left(\frac{M_{\rm h}}{M_{1c}}\right)^{-\beta_{\rm c}} + \left(\frac{M_{\rm h}}{M_{1c}}\right)^{\gamma_{\rm c}}\right]^{-1}, \tag{A3}$$

where $(\frac{m_{\rm c}}{M})_0$ is the overall normalization, $\beta_{\rm c}$ and $\gamma_{\rm c}$ control $m_{\rm c}(M_{\rm h})$ at the low- and high-halo-mass end, respectively, and $M_{1c}$ is the characteristic halo mass scale. A modified Schechter function is used to model the CSMF of satellites,

$$\Phi_{\rm sat}(m_*|M_{\rm h}) = \frac{\Phi_s^*}{m_s}\left(\frac{m_*}{m_s}\right)^{\alpha_s}\exp\left[-\left(\frac{m_*}{m_s}\right)^2\right], \tag{A4}$$

where $\alpha$ is the low-mass end slope,

$$\alpha = \alpha_0 + \alpha_s \times \log\left(\frac{M_{\rm h}}{{\rm M}_\odot}\right) \tag{A5}$$

$\Phi_s^*$ is the normalization,

$$\Phi_s^*(M_{\rm h}) = \Phi_0\left(\frac{M_{\rm h}}{{\rm M}_\odot}\right), \tag{A6}$$

and $m_s$ is the characteristic stellar mass in the distribution of satellites,

$$m_{\rm s}(M_{\rm h}) = 2M_{\rm h}\left(\frac{m_{\rm s}}{M}\right)_0\left[\left(\frac{M_{\rm h}}{M_{1s}}\right)^{-\beta_{\rm s}} + \left(\frac{M_{\rm h}}{M_{1s}}\right)^{\gamma_{\rm s}}\right]^{-1} \tag{A7}$$

which has the same functional form as $m_{\rm c}(M_{\rm h})$.

Equipped with the CSMF, we can calculate the abundance and clustering of galaxies. For example, the SMF can be derived as follows,

$$\Phi(m_*) = \int_0^\infty \Phi(m_*|M_{\rm h})n(M_{\rm h}){\rm d}M_{\rm h}, \tag{A8}$$

where $n(M_{\rm h})$ is the halo mass function (HMF). In this paper, we use the HMF from Tinker et al. (2008). The galaxy power spectrum as a function of $m_*$ is

$$P_{\rm gal}(k|m_*) = P_{\rm 1h}(k|m_*) + P_{\rm 2h}(k|m_*). \tag{A9}$$

The 1-halo term comes from galaxy pairs in the same halo,

$$P_{\rm 1h}(k|m_*) = \frac{1}{\Phi(m_*)^2}\int n(M_{\rm h})[\Phi_{\rm sat}(m_*|M_{\rm h})^2 u_{\rm g}(k|M_{\rm h})^2 + 2\Phi_{\rm cen}(m_*|M_{\rm h})\Phi_{\rm sat}(m|M_{\rm h})u_{\rm g}(k|M_{\rm h})]{\rm d}M_{\rm h}. \tag{A10}$$

Here $u_{\rm g}(k|M_{\rm h})$ is the normalized Fourier transform of the galaxy density distribution within a halo of mass $M_{\rm h}$, assumed to be an NFW profile (Navarro, Frenk & White 1997) truncated at the virial radius. The 2-halo term comes from galaxy pairs in separate haloes,

$$P_{\rm 2h}(k|m_*) = \left[\int {\rm d}M_{\rm h}n(M_{\rm h})b(M_{\rm h})\frac{\Phi(m_*|M_{\rm h})}{\bar{n}(m_*)}u_{\rm g}(k|M_{\rm h})\right]^2 P^{\rm lin}(k) \tag{A11}$$

Here $P^{\rm lin}(k)$ is the linear dark matter power spectrum, $b(M_{\rm h})$ is the bias factor as a function of $M_{\rm h}$ (Tinker et al. 2008). The projected correlation function at a given stellar mass is

$$w_p(r|m_*) = \int \frac{k{\rm d}k}{2\pi}P_{\rm gal}(k|m_*)J_0(kr), \tag{A12}$$

where $J_0(x) = \sin(x)/x$ is the zeroth-order Bessel function.

There are a total of 11 parameters in the CSMF of the local Universe. We make use of Markov chain Monte Carlo (MCMC) methods to derive the posterior PDF for all parameters by fitting to the observed abundance and clustering properties. Specifically, we use MCMC to minimize the reduced chi-squared

$$\chi_{\rm r}^2 = \frac{1}{N_\Phi}\Sigma_1^{N_\Phi}[(\Phi_{\rm CSMF}(m_*) - \Phi_{\rm obs}(m_*))/\sigma_\Phi]^2 + \frac{1}{N_{\rm s}}\Sigma_1^{N_{\rm s}}\frac{1}{N_{\rm r}}\Sigma_1^{N_{\rm r}}[(w_{p_{\rm CSMF}} - w_{p_{\rm obs}})/\sigma_{w_p}]^2, \tag{A13}$$

where $N_\Phi$ is the number of data points in the SMF, $N_{\rm r}$ is the number of data points in each projected correlation function and $N_{\rm s}$ is the total number of correlation functions. The best-fitting value and the standard deviation for each parameter are listed in Table A1. The correlation matrix of the parameters in the CSMF of the local Universe is shown in Table A2.

To add in the redshift evolution of the CSMF, we adopt the following parametrization to describe the evolving $m_*-M_{\rm h}$ relation, following Moster et al. (2010). The evolution in the characteristic halo mass scale is parametrized as

$$\log M_1(z) = (1+z)^\mu \times \log M_1|_{z=0}. \tag{A14}$$

**Table A1.** Parameters in the CSMF of the local Universe. The first four parameters describe the distribution of central galaxies as a function of $m_*$ at fixed $M_{\rm h}$, which is assumed to follow a log-normal distribution. The last seven parameters describe the distribution of satellites as a function of $m_*$ at fixed $M_{\rm h}$, which is assumed to follow a modified Schechter function.

| Parameter | Best fit | Error | Description |
|---|---|---|---|
| $\log M_{1c}$ | 11.70 | 0.49 | Characteristic halo mass in the $m_*/M_{\rm h}$ ratio |
| $(m_c/M)_0$ | 1.73 | 0.07 | Overall normalization |
| $\beta_c$ | 1.16 | 0.06 | Power-law slope of $m_*/M_{\rm h}$ at the low-mass end |
| $\gamma_c$ | 0.71 | 0.03 | Power-law slope of $m_*/M_{\rm h}$ at the high-mass end |
| $\log M_{1s}$ | 12.62 | 0.55 | Characteristic halo mass in the $m_*/M_{\rm h}$ ratio |
| $(m_s/M)_0$ | 2.32 | 0.13 | Overall normalization |
| $\beta_s$ | 2.38 | 0.33 | Power-law slope of $m_*/M_{\rm h}$ at the low-mass end |
| $\gamma_s$ | 0.97 | 0.05 | Power-law slope of $m_*/M_{\rm h}$ at the high-mass end |
| $-\log \Phi_0$ | 13.11 | 0.54 | Overall normalization in the number of satellites |
| $-\alpha_0$ | 0.28 | 0.11 | Power-law slope in $N_{\rm sat}$ |
| $-\alpha_s$ | 0.06 | 0.01 | Power-law slope in $N_{\rm sat}$ at the low-mass end |

And the overall normalization in the stellar-to-halo mass ratio is parametrized as

$$\left(\frac{m}{M}\right)_0(z) = (1+z)^{\nu} \times \left(\frac{m}{M}\right)_0|_{z=0} \tag{A15}$$

Finally, the power-law slope at the high-mass and low-mass end are parametrized as

$$\gamma(z) = (1+z)^{\gamma_1} \times \gamma|_{z=0}, \tag{A16}$$

and

$$\beta(z) = \beta|_{z=0} + \beta_1 \times z, \tag{A17}$$

respectively. We use the SMF in the high-$z$ Universe to constrain the redshift evolution of the $m_*-M_{\rm h}$ relation. The best-fitting value and the standard deviation for each parameter are listed in Table A3. The correlation matrix of the four parameters used to describe the redshift evolution of the CSMF is shown in Table A4. We have also tried to use eight evolution parameters to allow different redshift evolution for the central and satellite population. However, the parameters are highly correlated and the uncertainties on these parameters are very large from MCMC chains.

**Table A3.** The redshift evolution parameters in the CSMF.

| Parameter | Best fit | Error |
|---|---|---|
| $\mu$ | 0.028 | 0.010 |
| $\nu$ | 0.780 | 0.176 |
| $\beta_1$ | 0.079 | 0.133 |
| $\gamma_1$ | −0.061 | 0.268 |

**Table A4.** The correlation matrix of the redshift evolution parameters in the CSMF.

| | $\mu$ | $\nu$ | $\beta_1$ | $\gamma_1$ |
|---|---|---|---|---|
| $\mu$ | 1.00 | 0.57 | −0.64 | 0.76 |
| $\nu$ | | 1.00 | −0.02 | 0.85 |
| $\beta_1$ | | | 1.00 | −0.27 |
| $\gamma_1$ | | | | 1.00 |

## APPENDIX B: THE PROJECTED TWO-POINT CORRELATION FUNCTION

The spatial two-point correlation function is often used to study galaxy clustering. It is defined as the probability of finding a galaxy pair at a given separation, in excess of that in a random Poisson distribution. We use the Landy & Szalay (1993) estimator

$$\xi(r_p,\pi) = \frac{1}{\rm RR}\left[{\rm DD}\left(\frac{n_R}{n_D}\right)^2 - 2{\rm DR}\left(\frac{n_R}{n_D}\right) + {\rm RR}\right]. \tag{B1}$$

Here $r_p$ and $\pi$ are the separations perpendicular and parallel to the line of sight, $n_D$ and $n_R$ are the mean densities of the galaxy and random catalogues, respectively. DD($r$), DR($r$) and RR($r$) are numbers of weighted galaxy–galaxy pairs, galaxy–random pairs and random–random pairs at separation $r$, respectively. For volume-limited samples, the weight applied to each galaxy is 1. When generating random catalogues for clustering calculation, the angular distribution of random galaxies is modulated by an angular

**Table A2.** The correlation matrix of the parameters describing the CSMF of the local Universe.

| | $\log M_{1c}$ | $(m_c/M)_0$ | $\beta_c$ | $\gamma_c$ | $\log M_{1s}$ | $(m_s/M)_0$ | $\beta_s$ | $\gamma_s$ | $-\log\Phi_0$ | $\alpha_0$ | $\alpha_s$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\log M_{1c}$ | 1.00 | −0.66 | −0.70 | 0.79 | −0.54 | 0.56 | −0.04 | −0.44 | −0.74 | −0.15 | 0.11 |
| $(m_c/M)_0$ | | 1.00 | 0.44 | −0.11 | 0.43 | −0.48 | 0.06 | 0.17 | 0.59 | 0.19 | −0.16 |
| $\beta_c$ | | | 1.00 | −0.55 | 0.13 | −0.15 | −0.12 | 0.11 | 0.48 | −0.11 | 0.15 |
| $\gamma_c$ | | | | 1.00 | −0.38 | 0.37 | −0.01 | −0.43 | −0.50 | −0.06 | 0.03 |
| $\log M_{1s}$ | | | | | 1.00 | −0.99 | −0.02 | 0.81 | 0.62 | 0.11 | −0.07 |
| $(m_s/M)_0$ | | | | | | 1.00 | −0.04 | −0.72 | −0.62 | −0.13 | 0.09 |
| $\beta_s$ | | | | | | | 1.00 | −0.18 | −0.01 | 0.21 | −0.21 |
| $\gamma_s$ | | | | | | | | 1.00 | 0.58 | 0.05 | −0.01 |
| $-\log\Phi_0$ | | | | | | | | | 1.00 | 0.17 | −0.09 |
| $\alpha_0$ | | | | | | | | | | 1.00 | −0.99 |
| $\alpha_s$ | | | | | | | | | | | 1.00 |

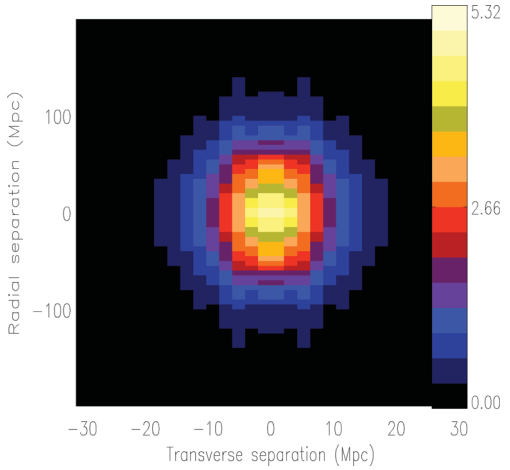


**Figure B1.** The redshift-space correlation function $\xi(r_p, \pi)$. The data from the first quadrant are repeated with reflection in both axes. The signal along the radial direction has been smoothed by a box filter of length 20 Mpc.

mask, which is generated using the optical flags to take into account the selection effect. In Fig. B1, we plot $\xi(r_p, \pi)$ of galaxies with $m_* > 10^{9.8}\,\mathrm{M}_\odot$ in the redshift bin $z1 = [0.2, 0.5]$, averaged over COSMOS and EGS. The signal from the first quadrant is repeated with reflection in both axes. In the absence of peculiar velocity and redshift error, $\xi(r_p, \pi)$ should be isotropic. The elongation of the signal along $\pi$ leads to a reduction in the clustering amplitude. The problem can be overcome by integrating $\xi(r_p, \pi)$ along $\pi$ to derive the projected correlation function,

$$w_p(r_p) = 2\int_0^\infty \xi(r_p, \pi)\mathrm{d}\pi. \tag{B2}$$

Fig. B1 also indicates that integrating $\xi(r_p, \pi)$ out to $\pi = 160$ Mpc should capture all correlated signal.

This paper has been typeset from a TEX/LATEX file prepared by the author.